Towards Feasible Instructor Intervention in MOOC Discussion Forums

Indicate Submission Type: Research-in-Progress

Muthu Kumar Chandrasekaran
Dept. of Computer Science
School of Computing
National University of Singapore
Singapore
muthu.chandra@comp.nus.edu.sg

Kiruthika Ragupathi
Centre for Development of
Teaching and Learning
National University of Singapore
Singapore
kiruthika@nus.edu.sg

Min-Yen Kan
Dept. of Computer Science
School of Computing
National University of Singapore
Singapore
kanmy@comp.nus.edu.sg

Bernard C.Y. Tan
Dept. of Information Systems
School of Computing
National University of Singapore
Singapore
btan@comp.nus.edu.sg

Abstract

Massive Open Online Courses allow numerous people from around the world to have access to knowledge that they otherwise have not. However, high student-to-instructor ratio in MOOCs restricts instructors' ability to facilitate student learning by intervening in discussions forums, as they do in face-to-face classrooms. Instructors need automated guidance on when and how to intervene in discussion forums. Using a typology of pedagogical interventions derived from prior research, we annotate a large corpus of discussion forum contents to enable supervised machine learning to automatically identify interventions that promote student learning. Such machine learning models may allow building of dashboards to automatically prompt instructors on when and how to intervene in discussion forums. In the longer term, it may be possible to automate these interventions relieving instructors of this effort. Such automated approaches are essential for allowing good pedagogical practices to scale in the context of MOOC discussion forums.

Keywords: *MOOCs, Instructor intervention, discussion forum, crowdsourcing*

Introduction

Massive Open Online Courses (MOOCs) have gained a significant amount of interest in higher education, given their potential to bring education to a large demographic. However, many pedagogical issues surrounding MOOCs require re-investigation in light of the lopsided student-instructor ratio, including ones that investigate best practices for student and instructor communication.

Discussion forums act as the primary mode of communication and interaction among instructors and students in MOOCs. Instructors use forums to communicate about lectures, readings, assignments, and have even been known to use structured open-ended questions to encourage discussions. Students on the other hand, use forums to collaborate in a "process of evolution and development" so as to encourage critical thought, expansion of horizons and deepened understanding of the topic (Trufant, 2003). Thus, a

discussion forum can be a useful tool to stimulate a sense of community and engagement that is available in a face-to-face classroom. Cormier (2013) argues that knowledge in a MOOC is emergent and dependent on the interaction with others – instructors and peers. Thus, the MOOC discussion forum is a critical source of information for instructors to understand their class and how students are interacting with the content (Stephens-Martinez *et al.* 2014).

Online discussion forums have been extensively studied in the computer-supported collaborative learning (CSCL) literature (*e.g.*, Stump *et al.* 2013). Collaborative learning environments provide necessary scaffolding enabling learners to perform tasks or learn concepts easily to take them to the next developmental level (Vygotsky, 1978). Online collaborative learning interactions, where students express their views, seek help from peers and discuss assignments can enhance academic discourse, foster high-level cognition of concepts (Dennen, 2008). A well-run discussion forum provides a sense of community and engagement, and can act as a substitute for the peer support available in a physical classroom while a hands-off approach to facilitating discussion forums can be detrimental to learners (Coetzee *et al.* 2015).

With over hundreds of thousands of students enrolling in a MOOC, it is difficult for the large majority of students to have meaningful interaction with the instructor. Learners easily "feel isolated and/or neglected" and believe that their instructor and peers are "ignoring their contributions" to the conversation in these MOOCs (Dolan, 2014). Mere exposure to information does not lead to the assimilation of knowledge, and hence for the MOOC learning experience to be effective, the onus falls on the instructor to cultivate the necessary engagement and dedication that students need to invest in order to gain value from the educational experience. For example, when an instructor responds at too high a rate, students do not respond to their peers' questions; whereas an almost-absent instructor induces very little participation from the students as well (Ghosh and Kleinberg, 2013). Brinton *et al.* (2013) shows that, over several weeks of a MOOC, student participation volume continued to decline despite active participation from the instructors.

Research suggests instructor participation and presence in forums improve student learning (Brookfield and Preskill, 2012) while others indicate that instructor presence in forums can only hinder student learning (Andresen, 2009). However, a third group argues that instructor interventions when used in moderation can be beneficial (Guldberg and Pilkington, 2007; Heckman and Annabi, 2006). Most of the earlier work focus on social media forums (Backstrom *et al.*, 2013), where the size and dynamics of interaction is quite different from that of the MOOCs.

Unlike many other works, our research work takes an instructor-centric viewpoint: to best utilize the time of MOOC instructors, which student queries and posts need to be responded to? Our research progresses along two axes: (1) identifying the type of instructor responses (formally, "interventions") that support knowledge co-construction and its relationship to learning and knowledge transfer, and (2) automatically providing intervention prompts for instructors through a monitoring process using machine learning. To the best of our knowledge, studies on predicting instructor interventions in MOOC forums are only in their preliminary stages (Chaturvedi *et al.* 2014, Chandrasekaran *et al.* 2015) or use a coding framework to manually classify copious amounts of data into a manageable number of categories (Stump *et al.* 2013).

Theory

Many studies have confirmed the need for some degree of structuring to promote the collaborative learning processes by guiding the participants through appropriate interventions (Chaturvedi *et al.* 2014; De Wever *et al.* 2007; Deslauriers, 2011; Dillenbourg and Jermann, 2010; Hämäläinen and Häkkinen, 2010; Kirschner *et al.* 2008; Slof *et al.* 2010; Stegmann *et al.* 2007). For instructors to make informed decisions and provide rich, holistic interactive support to their students, they need to understand when and how to intervene in forum discussions.

In particular, at the heart of our research is the idea of transactivity (Berkowitz and Gibbs, 1983) – the process in which learners relate to, elaborate on, and refer to the thoughts and ideas expressed in earlier conversations so as to integrate and co-construct a shared knowledge/understanding based on the reasoning of their instructor and peers. In our work, we propose to further classify such transactive contributions by coding the text according to the ten types of transactive contributions. Joshi and Rosé (2007) have also used a version of this framework to identify transactive portions of students' dialogue as candidates for its summary.

An instructor's skillful facilitation of a threaded discussion can promote students' critical thought, where the instructor poses questions and challenges ("justification requests") that encourage students to clarify and explain their positions, as well as construct new knowledge (Trufant, 2003). Another skill is the weaving together of recently posted ideas salient to the discussion. Weaving summarizes the discussions and extracts its major themes and disagreements ("integration"), presents conflicting opinions ("reasoning critique"), brings in new material periodically to freshen up the discussion ("extension"), provokes and instigates controversy, replaces missing cues by contextualizing, thus establishing general topicality (Winograd, 2003).

As seen, the instructor bears the responsibility of facilitating and orchestrating the discussion to enhance student learning. Therefore, it is essential that we investigate (1) when instructors can intervene so that students benefit maximally from the discussions in the forum; and (2) how to support instructor intervention effectively in a MOOC environment. We can thus term such instructor interventions to the discussion as scaffoldings for transactivity (Sionti *et al.*, 2011). Hence, it is logical to ask: Do instructor interventions in MOOC forums provide such scaffolding? Do students learn less from discussions without such interventions? To answer these, we first delineate categories of peer interventions from those of instructor interventions (*cf* Table 1), and focus on instructor interventions – justification request, extension, reasoning critique, and integration/summing up – and how facilitation through these instructor interventions enhances students' learning experience.

Peer Interventions	**Instructor Interventions**
Feedback Request	Justification Request
Paraphrase	Extension
Juxtaposition	Reasoning Critique
Refinement	Integration / Summing up
Clarification	
Completion	

Table 1. Our proposed taxonomy of transactive contributions.

Methods

We employ supervised machine learning and statistical natural language processing (NLP) to analyze discussion forum text. To perform machine-aided content analysis, the discussion forum text needs to be annotated. Due to the diversity in the nature of interventions, participation volume and posts observed across different MOOCs (Brinton *et al.* 2013; Rossi *et al.* 2014; Chandrasekaran *et al.* 2015), there is a need to annotate MOOC data from diverse disciplines for our work to generalize. A recent work (Chandrasekaran *et al.* 2015) reports inability to replicate other published findings (Chaturvedi *et al.* 2014; Ramesh *et al.* 2014) and attributed this failure to the wide variation among MOOCs. As annotation efforts are costly in time and effort, we plan to limit human annotation efforts to a seed corpus. We propose to explore semi-supervised machine learning methods to tag unlabeled data to expand the corpus.

We take a two-pronged approach to analyse the nature and types of intervention across MOOCs. First, we propose probabilistic models such as conditional random fields (Lafferty *et al.* 2001) that can predict intervention to maximize learning gains while optimizing instructor time. The models would be evaluated objectively using statistics collected from the annotated corpus. Second, we will perform a qualitative

content analysis of the annotated corpus to identify discourse markers in the student–student and student–instructor dialogues to measure learning gains. We anticipate complementary findings to emerge from the two methods. We now describe our approach in detail.

Data

We have collected a large-scale, multi-purpose dataset of discussion forums from MOOCs – a total of 33,665 threads of which 10,035 (~30%) were intervened. **Error! Reference source not found.**2 shows more detailed statistics of this dataset. An important desideratum was to collect forum data from a wide variety of different types of courses spanning the full breadth of disciplines: sciences, humanities and engineering. We collected forum threads from 61 completed courses from the Coursera platform, amounting to roughly 8% of the full complement of courses that Coursera offers[1]. Coursera allows instructors to divide its forums into various sub-forums to easily organise discussions on different aspects of the course as well as on specific topics the instructor intends to focus. In general, we find that the number of interventions varies significantly across these sub-forums (see Figure) with questions on exams and course logistics receiving the most instructor interventions. We omit sub-forums on logistics from the corpus as they are not associated with learning. Students post the most on homework and quizzes/exams sub-forums, perhaps because completion of homework and quizzes lead to a completion certificate. Although discussion threads in weekly lecture and general sub-forums may not be directly related to credits or the course, students still participate and learn from them. Therefore, we include these sub-forums for further annotation and analysis.

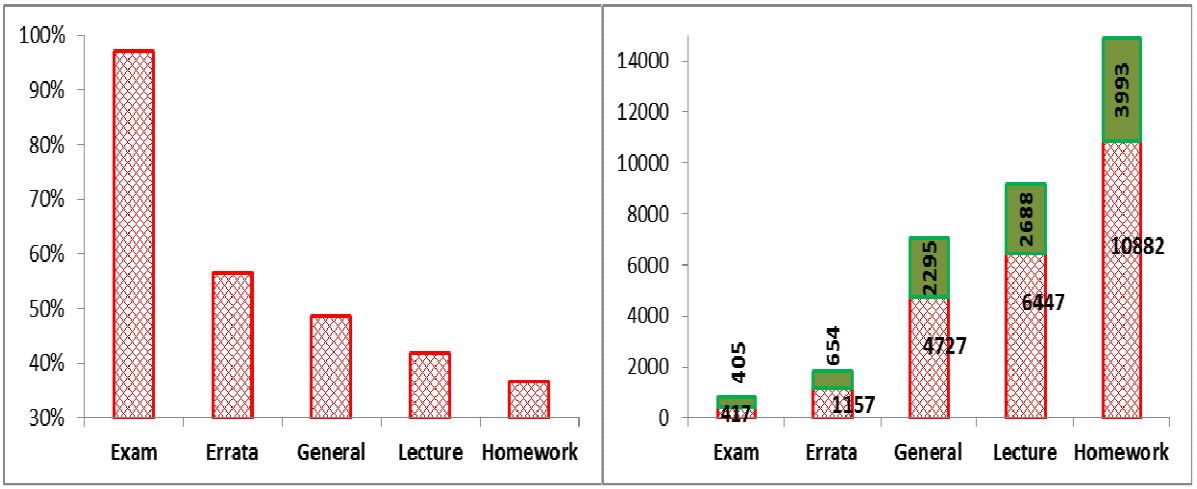

Figure 1a. *Intervention Ratio*, the ratio of number of threads intervened to those not intervened, across different forum types from 61 Coursera MOOCs. Figure 1b. Absolute numbers of intervened threads (solid green bar) and non-intervened threads (check patterned red bar).

Annotating a seed corpus

We divide our corpus of threads into three categories: those that have been intervened by an instructor or other staff, those that have been intervened but did not generate further discourse and those that were not intervened. We hypothesize student learning from these discussions will differ significantly. To identify the types of interventions that lead to learning, posts on intervened threads will need to be annotated further. We propose to annotate the corpus of intervened threads using crowdsourced human annotators (*i.e.,* global volunteers from Amazon Mechanical Turk[2]) and annotators recruited on-site. Preliminary

[1] As of December 2014, Coursera, a commercial MOOC platform: https://www.coursera.org, hosted 761 courses in English spanning 25 different subject areas

[2] https://www.mturk.com/mturk/welcome

results (Chandrasekaran *et al.* 2015) suggest that intervention may be subjective in nature. Therefore, to reliably annotate our corpus we will have three annotators to label each forum thread. Additionally, since significant domain knowledge is required to annotate forums well, we may observe that the quality of crowdsourced annotations may be insufficient for analysis. In this case, we may redundantly annotate them using on-site human annotators or by varying the number of annotators.

Crowdsourced annotation and evaluation has been used in MOOC studies before (Coetzee *et al. 2015;* Arguello *et al.* 2015). In Arguello *et al.'s* recent work, MOOC forum posts with speech acts were annotated by crowdsourced annotators using an inventory of six speech acts and an "others" class to tag a post on a MOOC forum thread. In our work, we seek to tag posts only by instructors using a tagset grounded on pedagogy theory and seek to identify only those interventions that signify student learning or lead to learning. Given the complete thread of posts from a MOOC forum up until an instructor intervenes, we ask annotators: (i) to link the instructor post to the earlier student post(s) to which it acts as a reply or comment (ii) to tag the instructor post with the most suitable type(s) from our predefined inventory of tags. We suppress posts after the instructor's first intervention to avoid annotators biasing towards some tags over others.

The procedure for onsite annotators is similar except that they are physically present at our venue and hence may receive additional training and help prior to annotation. Onsite annotators will include a mix of students and experts in education, pedagogy and education psychology. We propose to test an additional hypothesis with this setup: Do naïve annotators with training annotate as well as expert annotators on this task? This hypothesis seeks to estimate the cost of future annotation efforts, important for creating scalable corpora necessary for generalizable results from MOOC research.

Since annotation efforts are costly in time and effort, our project will be implemented in three phases. Each phase will involve annotation of incrementally larger samples of forum threads – Phase 1 (pilot phase), Phase 2 (from 3 MOOCs conducted by our university on the Coursera platform, hereafter referred to as internal MOOCs) and Phase 3 (final phase from MOOCs conducted by other Coursera partner universities, hereafter referred to as external MOOCs). We have obtained necessary approval from Coursera and our institutional review board to annotate forum text from internal MOOCs using both on-site and crowdsourced subjects. We have also procured session log data for internal MOOCs that enable measurement of student learning. The session log data from external MOOCs would be used on-site at Coursera to evaluate student learning for the final phase. Forum data from external MOOCs is to be procured through our MOOC Data Consortium (MOOCDC)[3], a community effort to create a standardised dataset for pooled evaluation of MOOC research studies. The MOOCDC is endorsed by Coursera which jointly holds the right to data generated from MOOCs on its platform. We call upon the research community members from institutions that partner with Coursera to join the consortium, contribute their institution's MOOC data and gain access to the larger collection.

Modeling

Using supervised sequence labeling techniques that take input from natural language dialogue and discourse analysis, we plan to learn patterns of intervention from the annotated data. Such models may predict optimal points during a forum discourse for the instructor to intervene in order to maximize student learning gains. They can also prioritize certain interventions over others, enabling better time management. We focus on modeling forum posts and its structure rather than topics and its evolution.

Recent studies on MOOC forums model posts in a thread by two ways: as discussion topics and as dialogues or speech entities. Stump *et al.* (2013) proposed a single taxonomy that accounts for such multidimensionality and allows a post to be tagged with more than one category. Ramesh *et al.*'s (2014) method uncovers post topics without prior annotations to a corpus. Ezen-Can and Boyer (2015) showed two different clustering methods that produced two clusters – one of *topics* and another of *dialogue acts*. Chaturvedi *et al.* (2013) automatically infers post categories from a corpus where every thread is tagged as "to be intervened" or "not intervened" and the inferred categories suggest a typical discourse leading to an intervention. However, all these studies on MOOC forums learn their topic and dialogue models from

[3] http://wing.comp.nus.edu.sg/downloads/moocdata

limited data from single MOOCs. Similar multi-dimensional models from conversational text in other domains exist. Ritter *et al.* (2010) and Backstrom *et al.* (2013) model the conversational structure in social media posts on Twitter and Facebook respectively. Ritter *et al.*'s model captures topical and dialogue aspects of Twitter conversations simultaneously.

We propose conditional random fields (CRF), a supervised machine learning model to learn sequence data, to model a MOOC discussion forum thread. CRFs have previously been used for sequence labelling NLP tasks such as tagging the words in a sentence with its parts-of-speech (Ratnaparkhi, 1996), word segmentation (Tseng *et al.*, 2005), foreground / background shadow segmentation (Wang *et al.*, 2006). Analogously, posts in a forum thread are also sequential with posts following earlier posts either as replies or as comments. They are often dictated by discourse patterns common in dialogues. For example, it is likely that a post posing a question is followed by an answer than by another question. Previously, Wang *et al.* (2012) had used CRF to model thread posts in troubleshooting forums as a sequence while detecting solvedness of a discussion thread. The discourse in a MOOC forum setting may be assumed to be similar in structure and function to tutorial dialogues between students and instructors or teaching assistants. Our proposed framework using transactivity is one such model. Different from earlier works on MOOC discussion forums, we propose categories and corpus annotation using theoretically-grounded findings in linguistics, pedagogy and cognitive sciences. These categories that are qualitative markers of learning may: i) complement the statistical model of intervention and ii) improve the statistical model's predictive power.

Measures

We will measure key variables – both quantitative and qualitative – for learner performance from Coursera's exported data. Earlier findings (Anderson *et al.* 2013) showed that forum participation correlates with learning outcomes observed through course assignments, quizzes and other similar evaluations. For us to assess whether higher or deeper level of learning and knowledge creation has occurred we need to test the level of learning within the MOOC forum environment. However, learning gains through discussions on the forum are difficult to observe directly at scale.

Hence, content in the annotated seed corpus will be analyzed manually to find discussions with discourse markers of learning. We seek those discourse markers that indicate *transactivity* that are in turn indicative of collaborative learning. This will help us distil intervention types that mindfully engage the learner, and not only lead to student learning but also enable students become facilitators of learning. That is, to ascertain: i) whether the instructor intervention promotes knowledge creation and ii) which types of interventions promote knowledge creation. We seek evidence in the form of high student interaction after an instructor's intervention. We plan to measure the frequency at which students are responding to the instructor's post. In particular, we may look for student posts that (1) seek to clarify the instructor's post by asking further questions ("clarification"); (2) elaborate their position to defend against the instructor's critique ("refinement"); (3) construct detailed and well-developed summary based on the instructor's response ("paraphrase"); (4) provide contrasting viewpoints ("juxtaposition") or (5) complete reasoning left incomplete in earlier student posts ("completion").

These findings can then be verified over a larger unannotated corpus by (i) testing for correlation between occurrences of transactive discourse and threads with intervention, and (ii) testing for correlation between occurrences of transactive discourse and students with higher learning outcomes from the course.

Conclusion

Peer learning aided by instructor feedback has been shown to foster learning better than traditional lecture based teaching (Deslauriers, 2011). Without any form of triage, providing such crucial feedback can be infeasible in MOOCs due to the large scale of MOOC discussion forums. This calls for development of automated methods to aid instructors and their staff, such as teaching assistants, to efficiently utilize their bandwidth to respond and satisfy a maximal number of students. Such optimal and timely response

may encourage more students to participate in forums and reduce attrition or dropouts, a perennial problem in MOOCs.

We have proposed to study interventions and its types in MOOC forums. Since our typology is grounded on a pedagogical framework that aims to capture student learning that emerges during discussions, we hope to isolate interventions that signal learning. Such information paves way for strategic instructor intervention with two key benefits: maximal peer learning and efficient instructor bandwidth use.

Further, MOOC forum dashboards for instructors can be designed to display forum threads ranked by an automated triage. This may help instructors prioritize their time for optimal effective intervention. With such ranking in place, we may also decide to push notifications to the instruction staff for urgent cases, while leaving the less urgent ones for later viewing.

Once a MOOC dashboard is established, instructors would then be able to attend to threads that require their urgent attention first. We would then be able to gauge student learning by looking into the frequency at which students engage in deeper and higher order thinking, as also how they apply the new knowledge shared in the forum. We plan to examine these topics in a future study. Such a study would also measure the level of student participation after an instructor intervention, and the intervention types that student engage in when posting in forums. We also plan to assess the frequency of peer intervention types by students and the frequency of students' crossing over to post comments that are of instructor intervention types as shown in Table 1.

Acknowledgements

This research is funded in part by NUS Learning Innovation Fund – Technology grant # C-252-000-123-001. We immensely thank Coursera and its staff for their support of our research by working with us to clear legal barriers to data sharing among its partner institutes. We would especially like to thank Dr. Andreina Parisi-Amon and Mr. Mark Pan from Coursera for their support to our efforts towards convening the MOOC Data Consortium.

References

Arguello, J., & Shaffer, K. 2015. "Predicting Speech Acts in MOOC Forum Posts". To appear in P*roceedings of the 9th international conference on web and social media 2015 (ICWSM 15), Oxford, UK.*

Andresen, M. A. 2009. "Asynchronous discussion forums: Success factors, outcomes, assessments, and limitations". *Educational Technology and Society*, 12(1), pp. 249-257.

Backstrom, L., Kleinberg, J., Lee, L. and Danescu-Niculescu-Mizil, C. 2013. "Characterizing and curating conversation threads: expansion, focus, volume, re-entry". In *Proceedings of the 6th ACM international conference on Web search and data mining (WSDM), Rome, Italy,* pp. 13-22. ACM.

Berkowitz, M., & Gibbs, J. (1983) "Measuring the developmental features of moral discussion." Merrill-Palmer Quarterly, 29, pp. 399-410.

Brinton, C., Chiang, M., Jain, S., Lam, H., Liu, Z., and Wong, F. 2013. "Learning about social learning in MOOCs: From statistical analysis to generative model." In *IEEE transactions on Learning Technologies*. 7(4), pp. 346-359. IEEE.

Brookfield, S. D., and Preskill, S. 2012. "*Discussion as a way of teaching: Tools and techniques for democratic classrooms"* (2nd ed.). San Francisco, CA: Jossey-Bass.

Chandrasekaran, M. K., Kan, M.-Y., Ragupathi, K., Tan, B. C. Y. 2015. "Learning instructor intervention from MOOC forums". In *Proceedings of the 8th International Conference on Educational Data Mining, Madrid, Spain*. pp. 218-225. International Education Data Mining Society.

Chaturvedi, S., Goldwasser, D., Daume, H., 2014, "Predicting instructor intervention in MOOC forums". In *Proceedings of the 52nd Annual Meeting of the Association for Computational Lingusitics, Baltimore, MD, USA*. ACL.

Coetzee, D., Lim, S., Fox, A., Hartmann, B., and Hearst, M.A., 2015. "Structuring Interactions for Large-Scale Synchronous Peer Learning". In *Proceedings of the 18th ACM Conference on Computer-*

Supported Cooperative Work and Social Computing (CSCW), Vancouver, Canada. pp. 1139-1152 ACM.

Cormier, D. 2013. "Forget the learners, how do I measure a MOOC quality experience for ME! By Dave Cormier". MOOC Quality Project.

Dolan, V. 2014. "Massive online obsessive compulsion: What are they saying out there about the latest phenomenon in higher education?". *The International Review of Research in Open and Distributed Learning*. 15(2), pp. 268-281.

Deslauriers, L., Schelew, E., and Wieman, C. (2011). "Improved learning in a large-enrollment physics class". *Science*, *332* (6031), pp. 862-864.

De Wever, B., Van Keer, H., Schellens, T., and Valcke, M. 2007. "Applying multilevel modelling on content analysis data: Methodological issues in the study of the impact of role assignment in asynchronous discussion groups". *Learning and Instruction*, 17, pp. 436-447.

Dennen, V. P. 2008, Pedagogical lurking: "Student engagement in non-posting discussion behavior". Computers in Human Behavior 24, 4, pp. 1624-1633.

Dillenbourg, P., and Jermann, P. 2010. "Technology for classroom orchestration". In *New science of learning,* pp. 525-552. Springer New York.

Ezen-Can, A. and Boyer, K. E. 2015. "Understanding student language: An unsupervised dialogue act classification approach". *Journal of Educational Data Mining (JEDM)*, 7(1), pp. 51–78.

Ghosh, A. and Kleinberg, J. 2013. "Incentivizing participation in online forums for education". In P*roceedings of 14th ACM Conference on Electronic Commerce, Philadelphia, PA, USA*.

Guldberg, K., and Pilkington, R. M. 2007. "Tutor roles in facilitating reflection on practice through online discussion". *Educational Technology and Society*, 10(1), pp. 61-72.

Heckman, R. and Annabi, H. 2006. "Cultivating voluntary online learning communities in blended environments". *Journal of Asynchronous Learning Networks*, 10(4), pp. 51-66.

Hämäläinen, R., and Häkkinen, P. 2010. "Teachers' instructional planning for computer-supported collaborative learning: Macro-scripts as a pedagogical method to facilitate collaborative learning". *Teaching and Teacher Education*, 26(4), pp. 871-877.

Joshi, M. and Rosé, C. P. 2007. "Using Transactivity in Conversation Summarization in Educational Dialog". In *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*, *Farmington, PA, USA*, pp. 53-56. Carnegie Mellon University and ISCA Archive.

Kirschner, P. A., Beers, P. J., Boshuizen, H. P. A., and Gijselaers, W. H. 2008. "Coercing shared knowledge in collaborative learning environments". Computers in Human Behavior, 24, pp. 403-420.

Lafferty, J., McCallum, A., and Pereira, F. C. 2001. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, *Williamstown, MA, USA*, 282-289.

Ratnaparkhi, A. 1996. "A maximum entropy model for part-of-speech tagging". *In Proceedings of the first conference on empirical methods in natural language processing (EMNLP),* 1, pp. 133-142. ACL.

Ritter, A., Cherry, C., Dolan, B. 2010. "Unsupervised Modeling of Twitter Conversations." In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 172-180. ACL.

Sionti, M., Ai, H., Rosé, C. P., and Resnick, L. 2011. "A Framework for Analyzing Development of Argumentation through Classroom Discussions", In Niels Pinkwart & Bruce McClaren (Eds.) *Educational Technologies for Teaching Argumentation Skills*, Bentham Science.

Slof, B., Erkens, G., Kirschner, P. A., Jaspers, J. G. M., and Janssen, J. 2010. "Guiding students' online complex learning-task behavior through representational scripting". *Computers in Human Behavior*, 26 (5), pp. 927-939.

Stegmann, K., Weinberger, A., and Fischer, F. 2007. "Facilitating argumentative knowledge construction with computer-supported collaboration scripts". *International Journal of Computer-Supported Collaborative Learning*, 2(4), pp. 421-447.

Stephens-Martinez, K., Hearst, M.A., and Fox, A. 2014. "Monitoring MOOCs: Which Information Sources Do Instructors Value?", In *Proceedings of the first ACM Conference on Learning@ Scale, Vancouver, Canada*. pp. 79-88. ACM.

Stump, G. S., DeBoer, J., Whittinghill, J., and Breslow, L. 2013. "Development of a framework to classify MOOC discussion forum posts: Methodology and challenges". *In NIPS Workshop on Data Driven Education, Lake Tahoe, NV, USA*.

Trufant, L. W., 2003. "Move over Socrates: Online Discussion is here". *EDUCAUSE*. Retrieved from http://net.educause.edu/ir/library/pdf/ncp0330.pdf

Tseng, H., Chang, P., Andrew, G., Jurafsky, D., and Manning, C. (October, 2005). "A conditional random field word segmenter for sighan bakeoff 2005". In *Proceedings of the fourth SIGHAN workshop on Chinese language Processing,* 171.

Vygotsky, L. S. (1978). "Mind in society: The development of higher psychological processes". Cambridge, MA: Harvard University.

Wang, L., Kim, S. N., and Baldwin, T. 2012. "The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums". In *Proceedings of the twenty-second International Conference on Computational Linguistics (COLING)*, pp. 2739-2756. ACL.

Wang, Y., Loe, K. F., and Wu, J. K. (2006). "A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(2), pp. 279-289.

Winograd, D. 2003. "The roles, functions and skills of moderators of online educational computer conferences for distance education". *Computers in the Schools*, 20(3), pp. 61-72.